Genome-wide association studies for genetic effects that change during lactation in dairy cattle

Lu, H., & Bovenhuis, H.

INTERPRETIVE SUMMARY

The present study aimed to identify QTL whose effects change during lactation using four different GWAS approaches. Twenty chromosomal regions were detected with effects on milk protein content, however, there was no evidence that their effects changed during lactation. Five chromosomal regions were detected whose effects on milk protein content changed during lactation, from which three were only identified based on GWAS for genotype by lactation stage interaction. Identification of QTL whose effects change during lactation are expected to help elucidate the genetic and biological background of milk production.

Genome-wide association studies for genetic effects that change during lactation in dairy cattle

Haibo Lu and Henk Bovenhuis[1]

Animal Breeding and Genomics, Wageningen University and Research, P.O. Box 338, 6700 AH, Wageningen, the Netherlands.

[1]Corresponding author: henk.bovenhuis@wur.nl

ABSTRACT

Genetic effects on milk production traits in dairy cattle might change during lactation. However, most genome-wide association studies (**GWAS**) for milk production traits assume that genetic effects are constant during lactation. This assumption might lead to missing these QTL whose effects change during lactation. This study aimed to screen the whole genome specifically for QTL whose effects change during lactation. For this purpose, four different GWAS approaches were performed using test-day milk protein content records: 1) separate GWAS for specific lactation stages; 2) GWAS for estimated Wilmink lactation curve parameters; 3) a GWAS using a repeatability model where SNP effects are assumed constant during lactation; and 4) a GWAS for genotype by lactation stage interaction using a repeatability model and accounting for changing genetic effects during lactation. Separate GWAS for specific lactation stages suggested that the detection power greatly differs between

lactation stages and that genetic effects of some QTL change during lactation. GWAS for estimated Wilmink lactation curve parameters detected many chromosomal regions for Wilmink parameter *a* (protein content level), whereas two regions for Wilmink parameter *b* (decrease in protein content towards nadir) and no regions for Wilmink parameter *c* (increase in protein content after nadir). Twenty chromosomal regions were detected with effects on milk protein content, however, there was no evidence that their effects changed during lactation. For five chromosomal regions located on chromosomes 3, 9, 10, 14, and 27 there was significant evidence for genotype by lactation stage interaction and thus that their effects on milk protein content changed during lactation. Three of these five regions were only identified using a GWAS for genotype by lactation stage interaction. Our study demonstrated that GWAS for genotype by lactation stage interaction offers new possibilities to identify QTL involved in milk protein content. The performed approaches can be applied to other milk production traits. Identification of QTL whose genetic effects change during lactation will help elucidate the genetic and biological background of milk production.

Key words: GWAS, genetic effect, longitudinal trait, genotype by lactation stage interaction

INTRODUCTION

Quantitative genetic studies have shown that the additive genetic variance for milk production traits changes during lactation (e.g. Jakobsen et al., 2002, Druet et al., 2005) and genetic correlations between milk production traits in early and late lactation differ from unity (e.g. Druet et al., 2003, Bastin et al., 2011). Furthermore, for the diacylglycerol O-acyltransferase 1 (***DGAT1***) K232A polymorphism it has been shown that its effect on milk production traits is not constant during lactation (e.g. Strucken et al., 2011, Szyda et al., 2014, Bovenhuis et al., 2015). In addition, results from gene expression studies show that the expression of several genes involved in milk production changes during lactation (e.g. Bionaz and Loor, 2011, Wickramasinghe et al., 2012). Therefore, genetic effects on milk production

traits might change during lactation. However, genome-wide association studies (**GWAS**) for milk production traits are mainly based on 305-day lactation records, which are summed or average test-day milk production records (e.g. Jiang et al., 2010, Cole et al., 2011). These studies detect QTL based on their average genetic effects during the whole lactation and assume that genetic effects of QTL related to milk production traits are constant. In a GWAS using models assuming constant genetic effects during lactation, QTL whose genetic effects change during lactation might not be detected (Lund et al., 2008, Ning et al., 2018).

Only a few studies specifically performed genome-wide screens for QTL whose genetic effects change during lactation (Strucken et al., 2012a, Macciotta et al., 2015). These GWAS were performed based on estimated lactation curve parameters or principal components and used relatively small data sets (less than 400 cows). Alternatively, screening the whole genome specifically for regions showing genotype by lactation stage interaction has not previously been carried out.

The objective of this study was to screen the whole genome for genetic effects that change during lactation. For this purpose we performed four GWAS approaches using test-day milk protein content in Dutch first parity Holstein cows: 1) separate GWAS for specific lactation stages; 2) GWAS for estimated Wilmink lactation curve parameters; 3) a GWAS using a repeatability model where SNP effects are assumed constant during lactation; and 4) a GWAS for genotype by lactation stage interaction using a repeatability model and accounting for changing genetic effects during lactation. This study will provide insight in differences between the four approaches and might lead to the detection of new QTL that would not have been detected when using models assuming genetic effects are constant. The results of this study are expected to further elucidate the genetic and biological background of milk protein content.

MATERIALS AND METHODS

Phenotypes and Genotypes

For this study, data on 1,829 Dutch Holstein first-parity cows were available. These cows are housed on 398 commercial herds in the Netherlands with at least three cows per herd. All cows were at least 87.5% Holstein-Friesian and descended from 5 proven bulls (98 to 196 daughters per sire), 50 test bulls (8 to 23 daughters per sire), and 15 other proven bulls (1 to 25 daughters per sire). Cows were milked twice daily and milk protein content was determined as part of routine milk recording using infrared spectroscopy (MilkoScan FT 6000, Foss Electric, Hillerød, Denmark) at the milk control station (Qlip, Zutphen, the Netherlands). The lactation was truncated at 390 days, each cow on average had 10.7 test-day records and the total number of test-day records was 19,593. Average milk protein content was 3.50% and the standard deviation was 0.31%.

DNA was isolated from blood samples and cows were genotyped using a customized 50k SNP chip (CRV, cooperative cattle improvement organization, Arnhem, the Netherlands) with the Infinium assay (Illumina, San Diego, CA). The SNP sequence were mapped using BLAST (http://www.ncbi.nlm.nih.gov/blast) and bovine genome assembly Btau 4.0 (Liu et al., 2009). In total, 1,868 cows were genotyped and 1,800 cows have both genotypes and test-day milk protein content records.

GWAS Approaches

If QTL effects change during lactation, separate GWAS for specific lactation stages might give different results. The GWAS signals might be strong during some parts of the lactation and weak or absent during other lactation stages. Therefore, in the first GWAS approach separate genome-wide associations were performed for specific lactation stages. For this purpose the lactation was divided in 26 lactation stages of 15 days each. Average number of test-day records for each lactation stage was 754. GWAS were performed based on data from

two consecutive lactation stage classes, e.g. lactation stages 1 & 2, 3 & 4 and so on. In this way most of the cows had at least one test-day record in each of the separate GWAS. Because the number of records per lactation stage decreased towards the end of lactation, data from lactation stages 21 to 26 were combined for the last GWAS. Combining lactation stage classes might in some cases result in multiple test day records per cow in a GWAS data set. In that case the first test day record of a cow was removed. The number of records in each lactation stage and each separate GWAS data set are shown in supplementary Table 1. The GWAS for specific lactation stages were performed using model [1]:

$$y_{jklmno} = \mu + b_1 * afc_{jklmno} + season_j + scode_k + lact_l + SNP_m + HTD_n + animal_o + e_{jklmno}, \quad [1]$$

where y_{jklmno} is test-day milk protein content; μ is the overall mean; afc_{jklmno} is a covariate describing the effect of age at first calving; $season_j$ is the fixed effect of calving season (June–August 2004, September–November 2004, and December 2004–February 2005); $scode_k$ is the fixed effect accounting for possible differences in genetic level between daughters of proven bulls, test bulls, and other proven bulls; $lact_l$ is the fixed effect of lactation stage (26 classes of 15 days each); SNP_m was the fixed effect of SNP genotype, modeled as a class variable; HTD_n was the random effect of herd-test-day, which was assumed to be distributed as $N(0, \boldsymbol{I}\sigma^2_{HTD})$; $animal_o$ was the random additive genetic effect of the individual and was assumed to be distributed as $N(0, \boldsymbol{A}\sigma^2_a)$ and e_{jklmno} was the random residual and was assumed to be distributed as $N(0, \boldsymbol{I}\sigma^2_e)$. **A** is the additive genetic relationships matrix constructed based on 14,062 animals and **I** is the identity matrix. Pedigree of the animals was traced back to five generations and provided by the Dutch herdbook (CRV, Arnhem, The Netherlands). Model [1] accounts for a lactation stage effect ($lact_l$) because each separate GWAS analyzed test-day records from at least two different lactation stage classes.

GWAS based on estimated lactation curve parameters were performed (Strucken et al., 2012a). In order to be able to compare our results with these GWAS, we performed the second GWAS approach. In these analyses, we first fitted a Wilmink lactation curve (Wilmink, 1987) to the test-day records of each cow using following model:

$$y_i = a + b * \exp^{-0.05*DIMi} + c * DIM_i + e_i, \quad [2]$$

where y_i is test-day milk protein content; DIM_i is days in milk; parameter a represents the milk protein content level; parameter b represents the decrease in protein content towards nadir; and parameter c represents the increase in protein content after nadir. Lactation curve parameters were estimated using the Procedure NLIN in SAS (SAS Inc., 1999). Subsequently GWAS for estimated lactation curve parameters, as proposed by Strucken et al. (2012a), were performed using the following model:

$$y_{jkmno} = \mu + b_1 * afc_{jkmno} + season_j + scode_k + SNP_m + HTD_n + animal_o + e_{jkmno}, \quad [3]$$

where y_{jkmno} are estimated lactation curve parameters a, b or c and the other model terms are as described for model [1].

A GWAS using a model that assumes that genetic effects are constant during lactation might not be able detect QTL whose genetic effects change during lactation (Lund et al., 2008, Ning et al., 2018). To investigate this hypothesis we performed a third GWAS approach using all test-day records and the following repeatability model that assumes that SNP effects are constant throughout the lactation:

$y_{jklmnop} = \mu + b_1 * afc_{jklmnop} + season_j + scode_k + lact_l + SNP_m + HTD_n + animal_o + pe_p + e_{jklmnop}$, [4]

where pe_p is the permanent environmental effect that was assumed to be distributed as $N(0, \boldsymbol{I}\sigma^2_{pe})$. Other model terms are as described for model [1] and lactation stage ($lact_l$) has 26 classes in this analysis.

Finally, we performed a fourth GWAS approach to specifically search for QTL whose effects change throughout lactation, i.e., SNP that show significant genotype by lactation stage interaction. For this purpose model [4] was extended with a SNP by lactation stage interaction term $(SNP*lact)_{lm}$**:**

$y_{jklmnop} = \mu + b_1 * afc_{jklmnop} + season_j + scode_k + lact_l + SNP_m + (SNP*lact)_{lm} + HTD_n + animal_o + pe_p + e_{jklmnop}$, [5]

where model terms are as described for model [1] and lactation stage ($lact_l$) has 26 classes in this analysis. For SNP that showed significant SNP by lactation stage interaction, the effects during the course of lactation were estimated using a model including the SNP by lactation stage interaction but without the main effects of SNP and lactation stage:

$y_{jklmnop} = \mu + b_1 * afc_{jklmnop} + season_j + scode_k + (SNP*lact)_{lm} + HTD_n + animal_o + pe_p + e_{jklmnop}$, [6]

where model terms are as described for model [1] and lactation stage class ($lact_l$) has 26 classes. A t-test was used to test the significance of the difference between any of two SNP genotypes within each lactation stage. If the P-value for the possible comparisons between any

of two SNP genotypes was smaller than 0.001, the SNP effect within that lactation stage was considered significant.

To test SNP by lactation stage interaction, any SNP genotype class in each lactation stage class needs to have a sufficiently large number of test-day records. SNP were not included in the GWAS if a genotype class contained less than 10 test-day records in any of the lactation stage classes. After this restriction, 30,348 SNP remained and the same SNP were used in the different GWAS approaches. All GWAS were performed in ASReml 4 (Gilmour et al., 2006).

Significance Threshold

The significance of SNP effects in GWAS approach 1 (separate lactation stages), GWAS approach 2 (Wilmink lactation curve parameters), GWAS approach 3 (repeatability model) and the SNP by lactation stage interaction effect in GWAS approach 4 were tested using the Wald F test statistic. Possible inflation of the test statistic was inspected based on quantile-quantile (**QQ**) plots where the observed $-\log_{10}$(P-value) was plotted against the expected $-\log_{10}$(P-value). The genome-wide significance threshold for the SNP effects was based on false discovery rate (**FDR**). FDR was calculated using the R package "qvalue" (Storey and Tibshirani, 2003) and FDR < 0.01 was considered significant. Previous GWAS for SNP by environment interaction observed a strong inflation of the test statistic for the interaction term (e.g. Voorman et al., 2011, Marigorta and Gibson, 2014). When the distribution of the test statistic under null hypothesis is unambiguous, permutation is a powerful strategy to estimate significance threshold (Churchill and Doerge, 1994, Doerge and Churchill, 1996). Therefore, the genome-wide significance threshold for the SNP by lactation stage interaction effect was not based on FDR but determined using permutation. In each permutation, all 30,348 SNPs of an animal were simultaneously assigned to a randomly selected other animal. Subsequently a GWAS was performed using the permuted genotypes. For each permutation the smallest genome-wide P-

value of the SNP by lactation stage interaction term was stored. Permutation was repeated 100 times to determine the 1% significance threshold for the interaction term.

RESULTS

The SNP with the highest -$\log_{10}$(P-value) for significant chromosomal regions (lead SNP) identified in the different GWAS approaches are in Table 1. Different chromosomal regions on the same chromosome are differentiated by letters.

Separate GWAS for Specific Lactation Stages

Manhattan plots of separate GWAS for specific lactation stages are shown in Figure 1. Results are presented for early lactation (lactation stages 1 & 2, Figure 1A), mid lactation (lactation stages 13 & 14, Figure 1B) and late lactation (lactation stages 21 to 26, Figure 1C). Manhattan plots of separate GWAS for other lactation stages are shown in supplementary Figure 1. Figure 1 and Table 1 demonstrate that there were large differences between lactation stages in number of detected chromosomal regions. In early lactation only one region on Bos taurus autosome (**BTA**) 6 significantly affected milk protein content. In mid lactation significant associations were detected on BTA 4, 5, 6, 10a, 10c, 14a, 15a, 20, 24, and 26. In late lactation significant associations were detected on BTA 6, 10b, 14a, and 16a. The region on BTA 6, which contains the casein gene cluster, was the only region that showed significant associations in all separate GWAS for specific lactation stages. The region on BTA 14a, which contains the *DGAT1*, did not show significant associations in early lactation and the significance of the GWAS signal showed large changes as lactation progressed (Table 1). Except BTA 6 and 14a, regions on BTA 4, 5, 10a, 10c, 15a, 20, 24, and 26 showed significant effects in mid lactation but no significant effects in early and late lactation. The region on BTA 10b and 16a showed significant associations in late lactation but no associations were detected in early and mid-lactation. These differences between lactation stages in number of detected chromosomal

regions and in their significance suggest that genetic effects of some QTL change during lactation.

GWAS for Wilmink Lactation Curve Parameters

Manhattan plots of GWAS for the three Wilmink lactation curve parameters are shown in Figure 2. For parameter *a*, representing the milk protein content level during lactation, significant SNPs were detected on BTA 1, 6, 8b, 9a, 14a, 15b, 16b, 20, 23, and 26 (Figure 2A). The strongest GWAS signals for parameter *a* were detected on BTA 6, 14a, and 20. For parameter *b*, which represents the decrease in protein content towards nadir, significant effects were detected on BTA 14a and 18 (Figure 2B). For parameter *c*, which represents the increase in protein content after nadir, no significant QTL were detected (Figure 2C).

GWAS Based on the Repeatability Model

The Manhattan plot for the GWAS using a repeatability model and assuming SNP effects are constant during lactation is shown in Figure 3. Significant chromosomal regions were detected on BTA 4, 6, 7, 8a, 10c, 11, 14a, 14b, 15a, 15b, 16a, 20 and 26. Strong GWAS signals were found on BTA 6, 14a, 15a, 15b, and 20; as 90% of the SNPs that passed the significance threshold were clustered in these chromosomal regions.

GWAS for SNP by Lactation Stage Interaction

The Wald F statistic for the SNP by lactation stage interaction effect showed a strong inflation, which is illustrated in the QQ plot (Supplementary Figure 2). To determine the appropriate threshold for the SNP by lactation stage interaction term permutation was performed. Based on 100 permutations the 1% genome-wide significance threshold was estimated to be $-\log_{10}$(P-value) = 18.6.

The Manhattan plot for the SNP by lactation stage interaction effect is shown in Figure 4. Significant SNP were detected on BTA 3, 9b, 10b, 14a, and 27. Estimated effects for the *(SNP*lact)* interaction term for the lead SNP in these chromosomal regions were obtained from

model [6]. Figure 5 shows the estimated effects of the lead SNP for the five regions that show significant SNP by lactation stage interaction. The lead SNP on BTA 14a showed a different pattern as compared to the lead SNP from the other significant regions. The lead SNP on BTA 3, 9b, 10b and 27 in general showed no significant effects in early and mid-lactation but SNP effects became significant towards late lactation whereas the lead SNP on BTA14a showed significant effects throughout the whole lactation except for early lactation (Figure 5).

Comparing Different GWAS Approaches

On BTA 8, 9, 10, 14, 15, 16, and 26, different GWAS approaches identified different lead SNP. A two-SNP analysis revealed that the lead SNP in region BTA 10b (at 46.6 Mbp and 48.7 Mbp, Table 1) were in strong linkage disequilibrium and they detected the same QTL. Similarly, the two lead SNP on BTA 26 were in strong linkage disequilibrium and represented the same QTL.

Some regions only showed significant effects in one of the GWAS approaches; BTA 5, 10a, and 24 only showed significant effects in the separate GWAS for specific lactation stages, BTA 9a and 16b were only significant for Wilmink parameter *a*, BTA 18 only showed significant effects for Wilmink parameter *b*, BTA 7, 8a, and 11 were only significant in the repeatability model [4] and BTA 3, 9b, and 27 only showed a significant SNP by lactation stage interaction effect. The region on BTA 14a showed highly significant effects in all GWAS approaches: all lactation stages except for lactation stages 1 & 2, Wilmink parameters *a* and *b*, the repeatability model and a highly significant SNP by lactation stage interaction.

Twenty chromosomal regions on BTA 1, 4, 5, 6, 7, 8a, 8b, 9a, 10a, 10c, 11, 14b, 15a, 15b, 16a, 16b, 20, 23, 24, and 26 did not show evidence for changing effect sizes during lactation: no clear pattern in the significance for different lactation stages, no significant effects for Wilmink parameters *b* and *c*, and no significant SNP by lactation stage interaction were detected. Five chromosomal regions on BTA 3, 9b, 10b, 14a and 27 showed significant SNP by lactation

stage interaction (model [5]), indicating that effects of these regions changed during lactation. BTA 10b was significant in the GWAS based on data from lactation stages 21 to 26 and also showed a strong but non-significant GWAS signal for Wilmink parameter *c* (Table 1, $-\log_{10}$(P-value) = 4.0). BTA 14a affected both the milk protein content level (Wilmink parameter *a*) and the shape of the lactation curve (Wilmink parameter *b*). BTA 3, 9b, and 27 showed significant SNP by lactation stage interaction but did not show significant effects in any of the other GWAS analyses we performed. These three chromosomal regions showed a clear increase in $-\log_{10}$(P-value) towards later lactation stages (Table 1, e.g. GWAS based on data from lactation stages 21 to 26, model [1]), although not significant. Furthermore, for Wilmink parameter *c*, the lead SNP on BTA 3 showed $-\log_{10}$(P-value) of 5.5, which is not significant at the applied threshold of FDR < 0.01 but significant at threshold of FDR < 0.05. BTA 9b showed a strong but non-significant GWAS signal for Wilmink parameter *c* (Table 1, $-\log_{10}$(P-value) = 5.1).

DISCUSSION

In this study we performed different GWAS using test-day milk protein content records. The objective was to specifically screen the genome for SNPs whose effects change during lactation. For this purpose four different approaches were performed: 1) separate GWAS for specific lactation stages; 2) GWAS for estimated Wilmink lactation curve parameters; 3) a GWAS using a repeatability model where SNP effects are assumed constant during lactation; and 4) a GWAS for genotype by lactation stage interaction using a repeatability model and accounting for genetic effects that change during lactation. Separate GWAS for specific lactation stages suggested that the detection power greatly differs between lactation stages and that effects of some QTL change during lactation. Many regions were detected for Wilmink parameter *a* whereas two regions were detected for Wilmink parameter *b* and no regions were detected for Wilmink parameter *c*. Twenty chromosomal regions were detected with effects on milk protein content, however, there was no evidence that their effects changed during lactation. A GWAS

specifically for SNP by lactation stage interaction identified five regions, from which three were not identified based on the other GWAS approaches we performed. To determine the appropriate significance threshold for the SNP by lactation stage interaction term permutation was used.

QTL for Milk Protein Content

In the current study regions on BTA 4, 6, 7, 8a, 10c, 11, 14a, 14b, 15a, 15b, 16a, 20 and 26 were identified using a repeatability model (model [4]) where SNP effects are assumed constant during lactation. Except for BTA 14a, we did not find evidence that effects of these regions changed during lactation, e.g. these regions were not significant for Wilmink parameters *b* or *c* and did not show significant SNP by lactation stage interaction. The region on BTA 6 contains the casein gene cluster (e.g. Ferretti et al., 1990, Threadgill and Womack, 1990) and the region on BTA 20 (35.9 Mbp, Table 1) is closed to the *Growth Hormone Receptor* (33.9 Mbp, Btau 4.0) gene (e.g. Arranz et al., 1998, Blott et al., 2003). These two QTL have been shown to have large effects on milk protein content. The region on BTA 10c (51.6 Mbp, Btau 4.0) was identical to the region detected by Schopen et al. (2011) in a GWAS for milk protein composition, which was based on largely the same animals and genotypes as used in the current study. On BTA 10 (46.6 Mbp, UMD 3.1) Nayeri et al. (2016) and Pausch et al. (2017) reported significant effects on milk protein content. Significant associations for chromosomal regions on BTA 4, 14b, 15a, 15b, and 16a are also in agreement with results from other GWAS (e.g. Buitenhuis et al., 2016, Pausch et al., 2017, Teissier et al., 2018). GWAS performed by Nayeri et al. (2016) and Pausch et al. (2017), which were based on large data sets, detected a number of chromosomal regions with effects on milk protein content that were not detected in the current study: regions on BTA 5, 29, and a second region on BTA 6. The reason we did not detect some of these regions might be related to power.

Regions on BTA 3, 9b, 10b, 14a, and 27 showed significant SNP by lactation stage interaction effects. The region on BTA 14a contains *DGAT1*, which has a major effect on several milk production traits (e.g. Grisart et al., 2002, Grisart et al., 2004, Bovenhuis et al., 2016). Effects of *DGAT1* on milk production traits change during lactation (Strucken et al., 2011, Szyda et al., 2014). Based on largely the same data as current study, Bovenhuis et al. (2015) described large *DGAT1* by lactation stage interaction on milk yield, fat content and protein content. Except for BTA 10b and 14a, the rest three regions were not significant in any of the other GWAS approaches we performed. However, these regions have been associated with milk production traits in other studies. Jiang et al. (2010) reported a QTL on BTA 3 (92.8 Mbp, Btau 4.0) with effects on milk and protein yield. Strucken et al. (2012a) reported significant effects for Wilmink parameters on BTA 3 (86.6, 115.9, and 116.9 Mbp) for milk protein yield. These GWAS signals are close to the region on BTA 3 (93.2 Mbp, Table 1) with significant SNP by lactation stage interaction. The region on BTA 27 (37.9 Mbp, Table 1) with a significant SNP by lactation stage interaction is closed to the *1-acylglycerol-3-phosphate O-acyltransferase 6* (***AGPAT6***) gene (38.9 Mbp, Btau 4.0). *AGPAT6* is involved in milk fat synthesis and has pleiotropic effects on other milk components (Littlejohn et al., 2014) and has been shown to affect milk fat yield and fat content over the first 60 days of lactation (Strucken et al., 2012b). Furthermore it has been shown that the expression of *AGPAT6* in the mammary gland increases over the first 60 days in lactation and decreases afterwards (Beigneux et al., 2006, Bionaz and Loor, 2008).

Approaches to Detect QTL whose Effects Change During Lactation

A simple approach to find indications for genetic effects that change during lactation is to split up the data and perform separate GWAS for different parts of the lactation. However, splitting up the data does not make optimal use of all available information and it does not provide a framework for significance testing of SNP whose genetic effect change during

lactation. Results from separate GWAS for different parts of the lactation show large differences in number of detected chromosomal regions: in early lactation only one region significantly affected milk protein content whereas in mid lactation up to ten different regions were detected. This shows that the power to detect QTL greatly differs between lactation stages. The difference in the number of QTL detected in the lactation stage and the change in additive genetic variance during lactation (Supplementary Table 1) also suggests that the effects of QTL change during lactation.

The low QTL detection power in early lactation as compared to later lactation stages can be explained by both a lower additive genetic variance and a higher residual variance: the heritability estimate for lactation stages 1 & 2 was 0.07 whereas for lactation stages 13 & 14 it was 0.63 (Supplementary Table 1). Separate GWAS for specific lactation stages is expected to be less powerful than GWAS based on the repeatability model as it uses approximately a ten times smaller number of records than the repeatability model. Counterintuitively, the results obtained from the GWAS based on the smaller data set from specific lactation stages (Model [1]) and based on the repeatability model using all test-day records (Model [4]) suggest that excluding test-day records from early lactation might be a means to increase the QTL detection power. For example, the $-\log_{10}$(P-value) for the region on BTA14a containing *DGAT1* based on the repeatability model [4] using all available test-day records was 33.1 whereas the GWAS for lactation stage 13 & 14 based on only 10% of the records, the $-\log_{10}$(P-value) for *DGAT1* reached 47.0 (Table 1). To check if excluding records can result in a stronger GWAS signal we performed an additional analysis using the repeatability model [4] but excluding data from lactation stages 1 to 4. This indeed increased significance of *DGAT1* from 33.1, based on all test day records, to 54.4 when analyzed based on a smaller data set consisting of test day records only from lactation stages 5 to 26. Difference between both homozygous *DGAT1* genotypes in lactation stage 1 &2 is -0.01 and in lactation stage 13 & 14 this is 0.26. In the repeatability

model [4] genotypic effects are averaged over the lactation and the difference between homozygous *DGAT1* genotypes is 0.18. As QTL detection power is directly related to QTL effect size these differences between *DGAT1* genotypes are part of the explanation why excluding test-day records from early lactation is a means to increase the QTL detection power.

To detect QTL whose effects change during lactation a two-step approach might be used where in a first analysis lactation curves are fitted to the test-day records and in a second analysis GWAS are performed based on estimated parameters. This approach has been used in other studies (e.g. Strucken et al., 2012a, Macciotta et al., 2015) and allows detection of QTL that affect the shape of the lactation curve. In our study these analyses mainly resulted in the detection of chromosomal regions that affected the milk protein content level (Wilmink parameter *a*) and only two chromosomal regions that affected the shape of the lactation curve (Wilmink parameters *b*) were detected. More subtle changes in the lactation curve, which were identified based on testing for SNP by lactation stage interaction, apparently are not picked up based on GWAS for Wilmink parameters. Using models that give a more accurate description of the lactation curve might be an alternative, however, these also require estimation of more parameters (e.g. Grossman and Koops, 2003).

The GWAS for Wilmink parameters detected several chromosomal regions affecting milk protein level (Wilmink parameter *a*), which were not detected in the repeatability model or in most of the lactation stage specific GWAS (regions on BTA 1, 8b, 9a, 16b, and 23). Therefore we concluded that these regions are likely false positives that might be a consequence of the two-step approach where differences in accuracies of estimated lactation curve parameters between cows are not taken into account in the GWAS. Consequently the obtained significance of SNP effects using this two-step approach are not correct and should be interpreted with caution.

A GWAS based on the repeatability model [4] assumes homogenous residual variance, which is an assumption that is violated in this study, especially in early lactation. To test the sensitivity of our results to heterogeneous residual variance we also performed a GWAS using phenotypes that were standardized based on the variance within each lactation stage class. This analysis did not result in the detection of other chromosomal regions than the ones reported in Table 1 (results not shown). The repeatability model assumes that SNP effects are constant throughout lactation and SNPs on BTA 4, 6, 7, 8a, 10c, 11, 14b, 15a, 15b, 16a, 20 and 26 seem to follow this assumption. The assumption of constant SNP effects might lead to missing time-dependent QTL effect (Lund et al., 2008, Ning et al., 2018). The effect of region on BTA 14a clearly changed during lactation but its effect still was detected due to its large average effect. SNPs on BTA 3, 9b, 10b, and 27, however, were not detected based on analyses using the repeatability model [4].

Testing for SNP by lactation stage interaction is an alternative approach to detect chromosomal regions whose effects change during lactation. A GWAS for SNP by lactation stage interaction identified three novel regions (BTA 3, 9b, and 27) that were not detected in other analyses. However, this model was not able to detect a region on BTA 16a, which showed a clear association in lactation stage 21 to 26 ($-\log_{10}$(P-value) = 7.0, Table 1). This illustrates that this approach is limited by the statistical power to detect interactions. In addition, determining the significance threshold for the interaction term needs permutation (test statistic inflation shown in supplementary Figure 2). To estimate significant threshold, we performed 100 permutations, which is computationally demanding.

Ning et al. (2018) used random regression to model changes in additive genetic, permanent environmental and SNP effects on test-day milk production records. Ning et al. (2018) concluded that the proposed model can control type I errors for QTL detection and has higher power as compared to a repeatability model. Theoretically random regression modeling also

would be suited for detecting QTL whose effects change during lactation. This would imply testing for the best polynomial fit of SNP effects might be computationally demanding.

Biological Interpretation

GWAS for SNP by lactation stage interaction identify regions whose genetic effects on milk protein content change during lactation. Effects on milk protein content can be due to effects on protein yield and milk yield. Change in genetic effects are in agreement with quantitative genetic studies that show that genetic variance and genetic correlations for milk production traits change, especially during the beginning and the end of lactation. Change of genetic effects are also confirmed based on gene expression studies (e.g. Bionaz and Loor, 2011, Wickramasinghe et al., 2012). Genetic effects of *DGAT1* on BTA 14a showed significant SNP by lactation stage interaction, which is mainly due to the lack of a *DGAT1* effect in early lactation (lactation stage 1 & 2, Figure 5D). The exact mechanism behind effects of *DGAT1* on milk protein synthesis remains unclear. Bovenhuis et al. (2015) indicated that most of the effects of *DGAT1* on milk production traits, like milk protein content, originated from the effect on water excretion (or dilution effect) and de novo FA synthesis. However, the *DGAT1* polymorphism also has significant effects on the yield of different milk proteins (Bovenhuis et al., 2016). In early lactation, dairy cows might suffer a negative energy balance. During this period after calving, dairy cows mobilize body reserves to balance the energy deficit due to the dramatic increase in milk yield and the restricted feed intake(e.g. Collard et al., 2000, Macciotta et al., 2015). Bovenhuis et al. (2015) suggested that in early lactation another DGAT enzyme, DGAT2 (Cases et al., 2001) might play a more important role than DGAT1 and this could be an explanation for the observed changes in DGAT1 effects which also might affect milk protein content.

Chromosomal regions on BTA 3, 9b, 10b, and 27 did not show significant effects on milk protein content in early and mid-lactation but only in late lactation (Figure 5). In late lactation,

most of the cows in our data were lactating and they were pregnant. However, because of different insemination and conception dates, dairy cows were in different pregnancy stages. Pregnancy has a negative effect on milk yield as a considerable amount of the nutrients are needed for the growth and maintenance of the developing fetus (e.g. Olori et al., 1997). Gestation stage also affects fat- and protein content of milk that increase as pregnancy advances (e.g. Olori et al., 1997). The mechanisms by which gestation affects milk yield and composition are mainly related to hormone-mediated partitioning of nutrients from milk production to pregnancy requirements. Furthermore, it is well established that the regulation of protein synthesis in the mammary gland is under control of hormones (Bionaz and Loor, 2011). Therefore, pregnancy might be a reason why genetic effects on milk protein content change during lactation, although the physiological mechanisms are still unknown. Associations between milk protein content and reproductive performance in dairy cows have been reported in several studies (e.g. Madouasse et al., 2010). It has been suggested that the association between milk protein content and reproductive performance is partly due to the negative energy balance in early lactation. Morton et al. (2016) indicated that factors determining milk protein content during the first 30 d of lactation are not identical to factors determining milk protein content in late lactation. Furthermore, Morton et al. (2016) suggested that milk protein content in late lactation is more important than milk protein content in early lactation for the milk protein content-reproductive performance relationship. This is in agreement with the hypothesis that pregnancy might be a reason why genetic effects on milk protein content change during lactation.

CONCLUSIONS

The current study aimed to detect genetic effects that change during lactation. For this purpose, four different GWAS approaches were performed for milk protein content. Separate GWAS for specific lactation stages suggested that the detection power greatly differs between

lactation stages and that genetic effects of some QTL change during lactation. GWAS for estimated Wilmink lactation curve parameters detected many QTL but these results should be interpreted with caution as they were based on a two-step approach. Twenty chromosomal regions were detected with effects on milk protein content, however, there was no evidence that their effects changed during lactation. Five chromosomal regions were detected whose effect on milk protein content change during lactation on BTA 3, 9b, 10b, 14a, and 27 , from which BTA 3, 9b, and 27 were only detected in GWAS for SNP by lactation stage interaction. The performed approaches can be used to other milk production traits. Exploring QTL whose effects change during lactation are expected to elucidate the genetic and biological background of milk production.

ACKNOWLEDGMENTS

Haibo Lu is financially supported by Sino-Dutch Dairy Development Centre (Beijing, China). Yachun Wang (China Agricultural University, Beijing, China) is acknowledged for assistance in project discussion. This study uses data generated as part of the Dutch Milk Genomics Initiative project, funded by Wageningen University and Research (Wageningen, the Netherlands), the Dutch Dairy Association NZO (Zoetermeer, the Netherlands), Cooperative Cattle Improvement Organization CRV (Arnhem, the Netherlands), and the Dutch Technology Foundation STW (Utrecht, the Netherlands).

REFERENCES

Arranz, J-J., W. Coppieters, P. Berzi, N. Cambisano, B. Grisart, L. Karim, F. Marcq, L. Moreau, C. Mezer, J. Riquet, P. Simon, P. Vanmanshoven, D. Wagenaar, and M. Georges. 1998. A QTL affecting milk yield and composition maps to bovine chromosome 20: a confirmation. Anim. Genet. 29:107-115.

Bastin, C., N. Gengler, and H. Soyeurt. 2011. Phenotypic and genetic variability of production traits and milk fatty acid contents across days in milk for Walloon Holstein first-parity cows. J. Dairy Sci. 94:4152-4163.

Beigneux, A. P., L. Vergnes, X. Qiao, S. Quatela, R. Davis, S. M. Watkins, R. A. Coleman, R. L. Walzem, M. Philips, K. Reue, and S. G. Young. 2006. Agpat6--a novel lipid biosynthetic gene required for triacylglycerol production in mammary epithelium. J. Lipid Res. 47:734-744.

Bionaz, M. and J. J. Loor. 2008. ACSL1, AGPAT6, FABP3, LPIN1, and SLC27A6 are the most abundant isoforms in bovine mammary tissue and their expression is affected by stage of lactation. J. Nutr. 138:1019-1024.

Bionaz, M. and J. J. Loor. 2011. Gene networks driving bovine mammary protein synthesis during the lactation cycle. Bioinform. Biol. Insights 5:83-98.

Blott, S., J. J. Kim, S. Moisio, A. Schmidt-Kuntzel, A. Cornet, P. Berzi, N. Cambisano, C. Ford, B. Grisart, D. Johnson, L. Karim, P. Simon, R. Snell, R. Spelman, J. Wong, J. Vilkki, M. Georges, F. Farnir, and W. Coppieters. 2003. Molecular dissection of a quantitative trait locus: A phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. Genetics 163:253-266.

Bovenhuis, H., M. H. P. W. Visker, N. A. Poulsen, J. Sehested, H. J. F. van Valenberg, J. A. M. van Arendonk, L. B. Larsen, and A. J. Buitenhuis. 2016. Effects of the diacylglycerol o-acyltransferase 1 (DGAT1) K232A polymorphism on fatty acid, protein, and mineral composition of dairy cattle milk. J. Dairy Sci. 99:3113-3123.

Bovenhuis, H., M. H. Visker, H. J. van Valenberg, A. J. Buitenhuis, and J. A. van Arendonk. 2015. Effects of the DGAT1 polymorphism on test-day milk production traits throughout lactation. J. Dairy Sci. 98:6572-6582.

Buitenhuis, B., N. A. Poulsen, G. Gebreyesus, and L. B. Larsen. 2016. Estimation of genetic parameters and detection of chromosomal regions affecting the major milk proteins and their post translational modifications in Danish Holstein and Danish Jersey cattle. BMC Genet. 17:114.

Cases, S., S. J. Stone, P. Zhou, E. Yen, B. Tow, K. D. Lardizabal, T. Voelker, and R. V. Farese Jr. 2001. Cloning of DGAT2, a Second Mammalian Diacylglycerol Acyltransferase, and Related Family Members. J. Biol. Chem. 276:38870-38876.

Churchill, G. A. and R. W. Doerge. 1994. Empirical threshold values for quantitative trait mapping. Genetics 138:963-971.

Cole, J. B., G. R. Wiggans, L. Ma, T. S. Sonstegard, T. J. Lawlor, Jr., B. A. Crooker, C. P. Van Tassell, J. Yang, S. Wang, L. K. Matukumalli, and Y. Da. 2011. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. BMC Genomics 12:408.

Collard, B. L., P. J. Boettcher, J. C. M. Dekkers, D. Petitclerc, and L. R. Schaeffer. 2000. Relationships Between Energy Balance and Health Traits of Dairy Cattle in Early Lactation. J. Dairy Sci. 83:2683-2690.

Doerge, R. W. and G. A. Churchill. 1996. Permutation tests for multiple loci affecting a quantitative character. Genetics 142:285-294.

Druet, T., F. Jaffrezic, D. Boichard, and V. Ducrocq. 2003. Modeling lactation curves and estimation of genetic parameters for first lactation test-day records of French Holstein cows. J. Dairy Sci. 86:2480-2490.

Druet, T., F. Jaffrezic, and V. Ducrocq. 2005. Estimation of genetic parameters for test day records of dairy traits in the first three lactations. Genetics, selection, evolution : GSE 37:257-271.

Ferretti, L., P. Leone, and V. Sgaramella. 1990. Long range restriction analysis of the bovine casein genes. Nucleic Acids Res. 18:6829-6833.

Gilmour, A. R., B. J. Gogel, B. R. Cullis, S. J. Welham, and R. Thompson. 2006. ASReml User Guide Release 1.0.

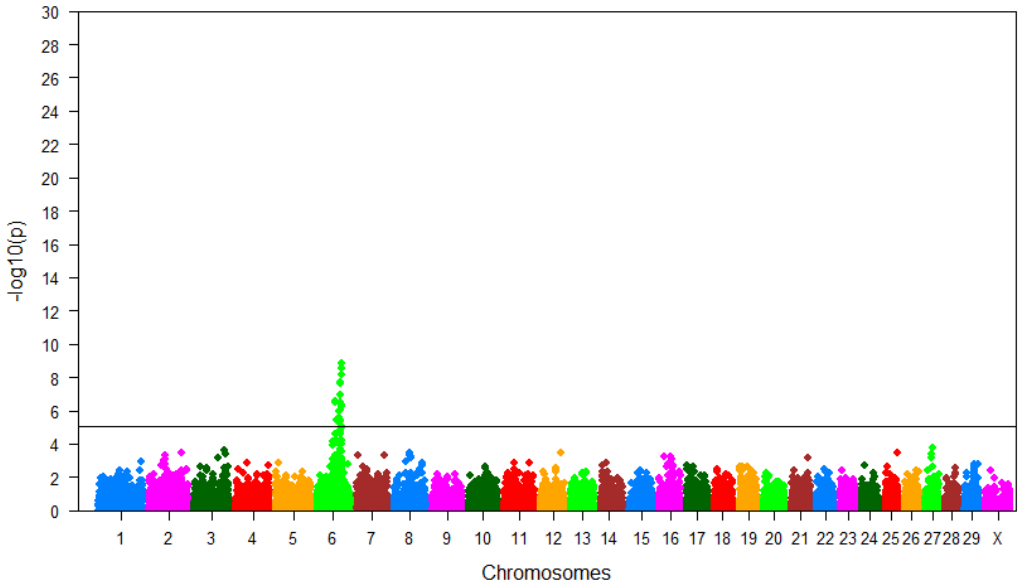
Grisart, B., W. Coppieters, F. Farnir, L. Karim, C. Ford, P. Berzi, N. Cambisano, M. Mni, S. Reid, P. Simon, R. Spelman, M. Georges, and R. Snell. 2002. Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. Genome Res. 12:222-231.

549 Grisart, B., F. Farnir, L. Karim, N. Cambisano, J. J. Kim, A. Kvasz, M. Mni, P. Simon, J. M.
550 Frere, W. Coppieters, and M. Georges. 2004. Genetic and functional confirmation of the
551 causality of the DGAT1 K232A quantitative trait nucleotide in affecting milk yield and
552 composition. Proc. Natl. Acad. Sci. U. S. A. 101:2398-2403.
553 Grossman, M. and W. J. Koops. 2003. Modeling extended lactation curves of dairy cattle: A
554 biological basis for the multiphasic approach. J. Dairy Sci. 86:988-998.
555 Jakobsen, J. H., P. Madsen, J. Jensen, J. Pedersen, L. G. Christensen, and D. A. Sorensen. 2002.
556 Genetic parameters for milk production and persistency for Danish Holsteins estimated in
557 random regression models using REML. J. Dairy Sci. 85:1607-1616.
558 Jiang, L., J. Liu, D. Sun, P. Ma, X. Ding, Y. Yu, and Q. Zhang. 2010. Genome wide association
559 studies for milk production traits in Chinese Holstein population. PLoS One 5:e13661.
560 Littlejohn, M. D., K. Tiplady, T. Lopdell, T. A. Law, A. Scott, C. Harland, R. Sherlock, K.
561 Henty, V. Obolonkin, K. Lehnert, A. Macgibbon, R. J. Spelman, S. R. Davis, and R. G.
562 Snell. 2014. Expression variants of the lipogenic AGPAT6 gene affect diverse milk
563 composition phenotypes in Bos taurus. PLoS One 9:e85757.
564 Liu, Y., X. Qin, X. Z. H. Song, H. Jiang, Y. Shen, K. J. Durbin, S. Lien, M. P. Kent, M. Sodeland,
565 Y. Ren, L. Zhang, E. Sodergren, P. Havlak, K. C. Worley, G. M. Weinstock, and R. A.
566 Gibbs. 2009. Bos taurus genome assembly. BMC Genomics 10:
567 Lund, M. S., P. Sorensen, P. Madsen, and F. Jaffrezic. 2008. Detection and modelling of time-
568 dependent QTL in animal populations. Genetics, selection, evolution : GSE 40:177-194.
569 Macciotta, N. P., G. Gaspa, L. Bomba, D. Vicario, C. Dimauro, M. Cellesi, and P. Ajmone-
570 Marsan. 2015. Genome-wide association analysis in Italian Simmental cows for lactation
571 curve traits using a low-density (7K) SNP panel. J. Dairy Sci. 98:8175-8185.
572 Madouasse, A., J. N. Huxley, W. J. Browne, A. J. Bradley, I. L. Dryden, and M. J. Green. 2010.
573 Use of individual cow milk recording data at the start of lactation to predict the calving to
574 conception interval. J. Dairy Sci. 93:4677-4690.
575 Marigorta, U. M. and G. Gibson. 2014. A simulation study of gene-by-environment interactions
576 in GWAS implies ample hidden effects. Front Genet 5:225.
577 Morton, J. M., M. J. Auldist, M. L. Douglas, and K. L. Macmillan. 2016. Associations between
578 milk protein concentration at various stages of lactation and reproductive performance in
579 dairy cows. J. Dairy Sci. 99:10044-10056.
580 Nayeri, S., M. Sargolzaei, M. K. Abo-Ismail, N. May, S. P. Miller, F. Schenkel, S. S. Moore,
581 and P. Stothard. 2016. Genome-wide association for milk production and female fertility
582 traits in Canadian dairy Holstein cattle. BMC Genet. 17:75.
583 Ning, C., D. Wang, X. Zheng, Q. Zhang, S. Zhang, R. Mrode, and J. F. Liu. 2018. Eigen
584 decomposition expedites longitudinal genome-wide association studies for milk production
585 traits in Chinese Holstein. Genetics Selection Evolution 50:
586 Olori, V. E., S. Brotherstone, W. G. Hill, and B. J. McGuirk. 1997. Effect of gestation stage on
587 milk yield and composition in Holstein Friesian dairy cattle. Livestock Production Science
588 52:167-176.
589 Pausch, H., R. Emmerling, B. Gredler-Grandl, R. Fries, H. D. Daetwyler, and M. E. Goddard.
590 2017. Meta-analysis of sequence-based association studies across three cattle breeds reveals
591 25 QTL for fat and protein percentages in milk at nucleotide resolution. BMC Genomics
592 18:
593 SAS Inc., Cary, NC. 1999. SAS Procedures Guide. Version 8.
594 Schopen, G. C. B., M. H. P. W. Visker, P. D. Koks, E. Mullaart, J. A. M. van Arendonk, and
595 H. Bovenhuis. 2011. Whole-genome association study for milk protein composition in dairy
596 cattle. J. Dairy Sci. 94:3148-3158.
597 Storey, J. D. and R. Tibshirani. 2003. Statistical significance for genomewide studies. Proc.
598 Natl. Acad. Sci. U. S. A. 100:9440-9445.

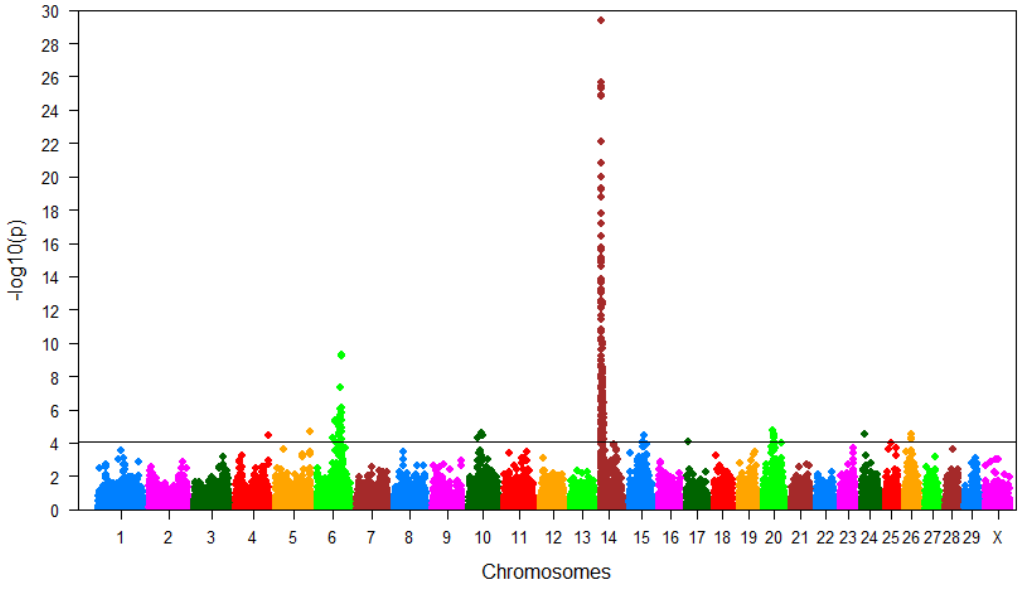
599 Strucken, E. M., R. H. Bortfeldt, D. J. de Koning, and G. A. Brockmann. 2012a. Genome-wide
600 associations for investigating time-dependent genetic effects for milk production traits in
601 dairy cattle. Anim. Genet. 43:375-382.
602 Strucken, E. M., R. H. Bortfeldt, J. Tetens, G. Thaller, and G. A. Brockmann. 2012b. Genetic
603 effects and correlations between production and fertility traits and their dependency on the
604 lactation-stage in Holstein Friesians. BMC Genet. 13:108.
605 Strucken, E. M., D. J. de Koning, S. A. Rahmatalla, and G. A. Brockmann. 2011. Lactation
606 curve models for estimating gene effects over a timeline. J. Dairy Sci. 94:442-449.
607 Szyda, J., J. Komisarek, and I. Antkowiak. 2014. Modelling effects of candidate genes on
608 complex traits as variables over time. Anim. Genet. 45:322-328.
609 Teissier, M., M. P. Sanchez, M. Boussaha, A. Barbat, C. Hoze, C. Robert-Granie, and P.
610 Croiseau. 2018. Use of meta-analyses and joint analyses to select variants in whole genome
611 sequences for genomic evaluation: An application in milk production of French dairy cattle
612 breeds. J. Dairy Sci. 101:3126-3139.
613 Threadgill, D. W. and J. E. Womack. 1990. Genomic analysis of the major bovine milk protein
614 genes. Nucleic Acids Res. 18:6935-6942.
615 Voorman, A., T. Lumley, B. McKnight, and K. Rice. 2011. Behavior of QQ-plots and genomic
616 control in studies of gene-environment interaction. PLoS One 6:e19416.
617 Wickramasinghe, S., G. Rincon, A. Islas-Trejo, and J. F. Medrano. 2012. Transcriptional
618 profiling of bovine milk using RNA sequencing. BMC Genomics 13:45.
619 Wilmink, J. B. M. 1987. Adjustment of Test-Day Milk, Fat and Protein Yield for Age, Season
620 and Stage of Lactation. Livestock Production Science 16:335-348.

TABLES AND FIGURES

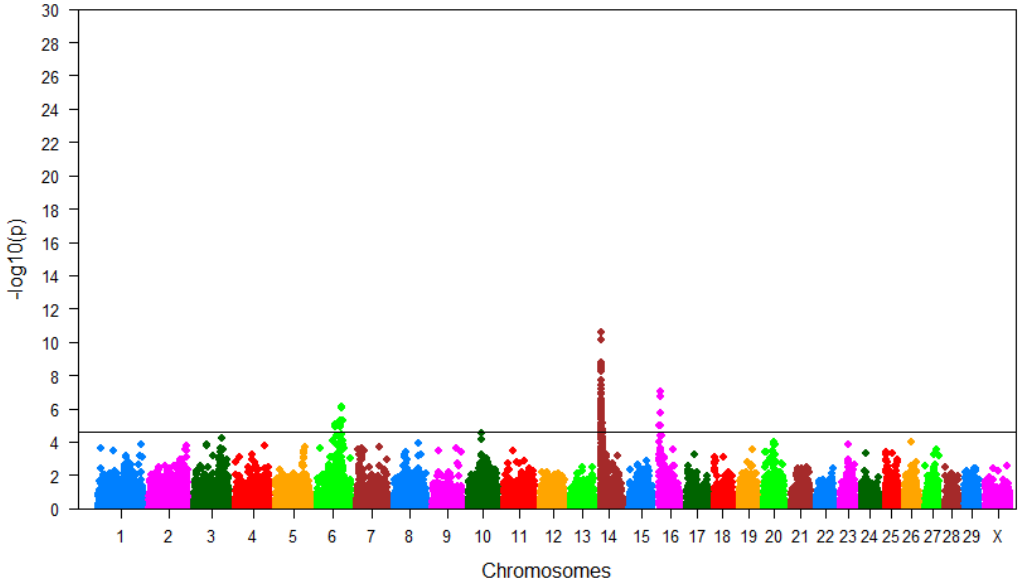
A



B



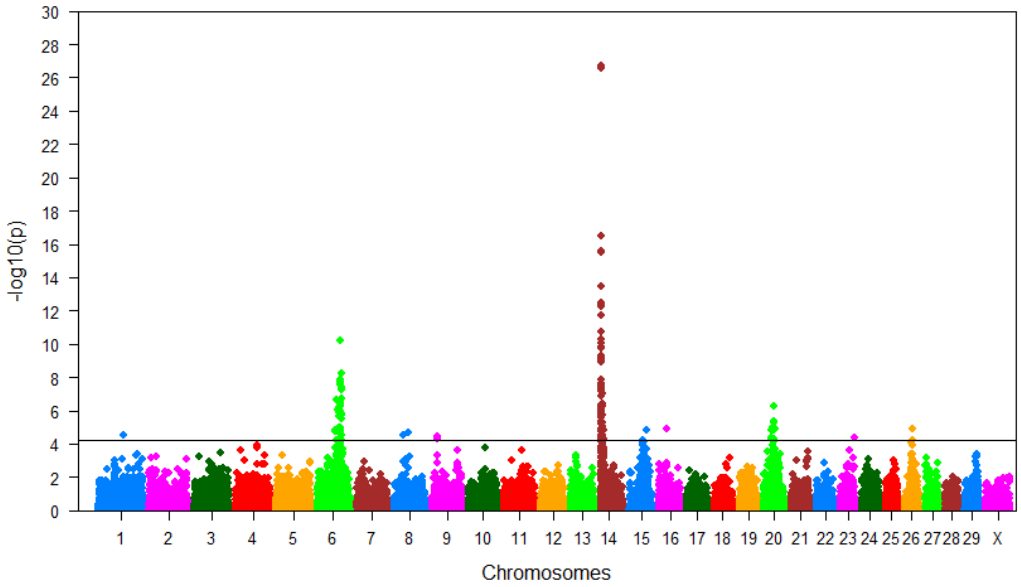
C



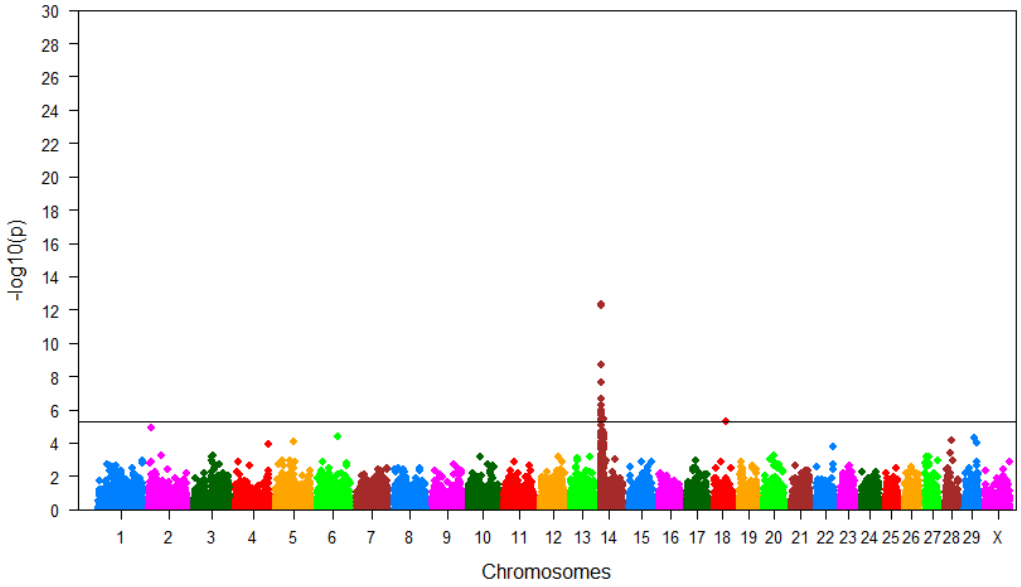
627

628 **Figure 1.** Manhattan plot for milk protein content in specific different lactation stages. A:
629 lactation stages 1 & 2 (day 0-30), B: lactation stages 13 & 14 (day 180-210) and C: lactation
630 stages 21 to 26 (day 300-390). The cut-off value for the y-axis is set at a $-\log_{10}$(P-value) of 30.
631 The horizontal line indicates a false discovery rate < 0.01.

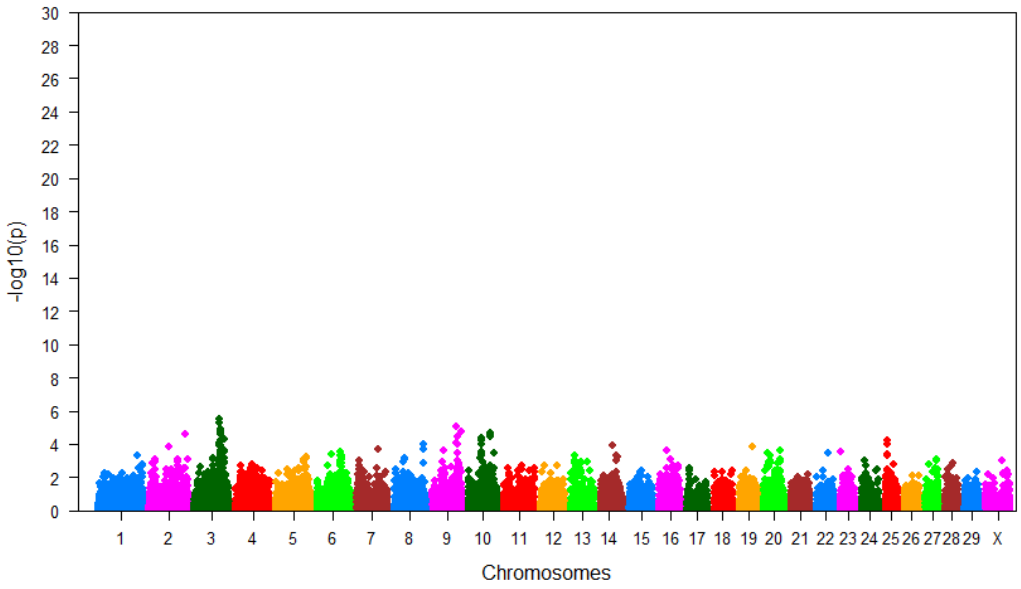
A


-log10(p)
Chromosomes
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 X

B

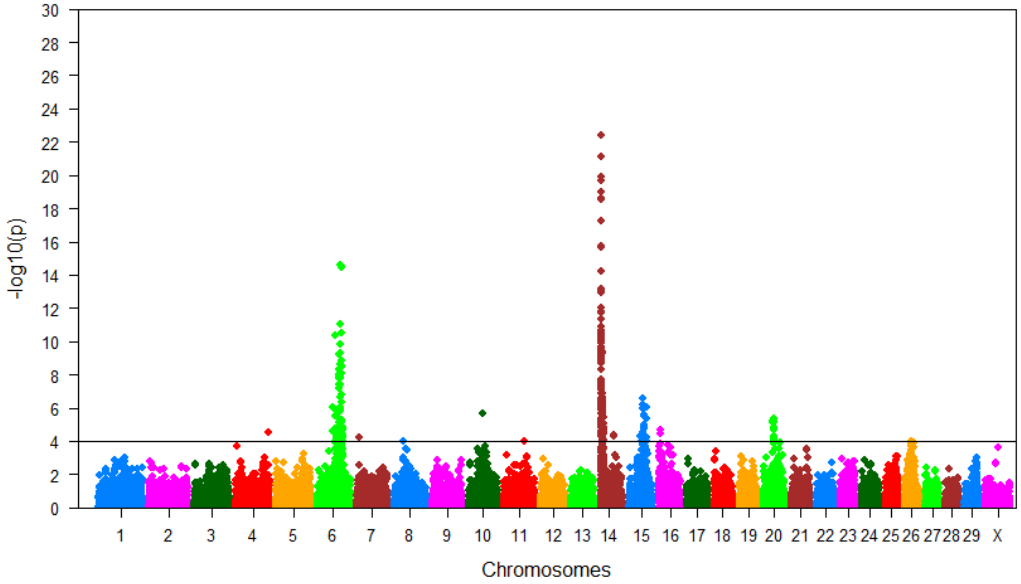

-log10(p)
Chromosomes
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 X

C



637

638 **Figure 2.** Manhattan plot for Wilmink lactation curve parameters fitted to milk protein test-
639 day records: A: Wilmink parameter *a*. B: Wilmink parameter *b*. C: Wilmink parameter *c*. The
640 cut-off value for the y-axis is set at a $-\log_{10}$(P-value) of 30. The horizontal line indicates a false
641 discovery rate < 0.01. In Figure 2C no threshold is indicated as none of the SNP effects were
642 significant.



643

644 **Figure 3.** Manhattan plot for milk protein content based on test-day milk protein content

645 records. The cut-off value for the y-axis is set at a $-\log_{10}$(P-value) of 30. The horizontal line

646 indicates a false discovery rate < 0.01.

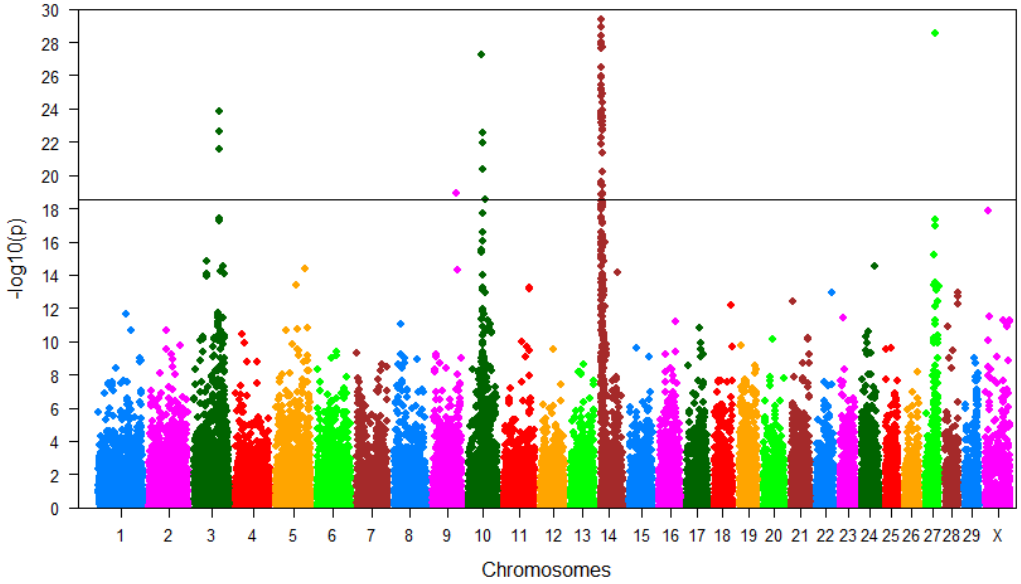
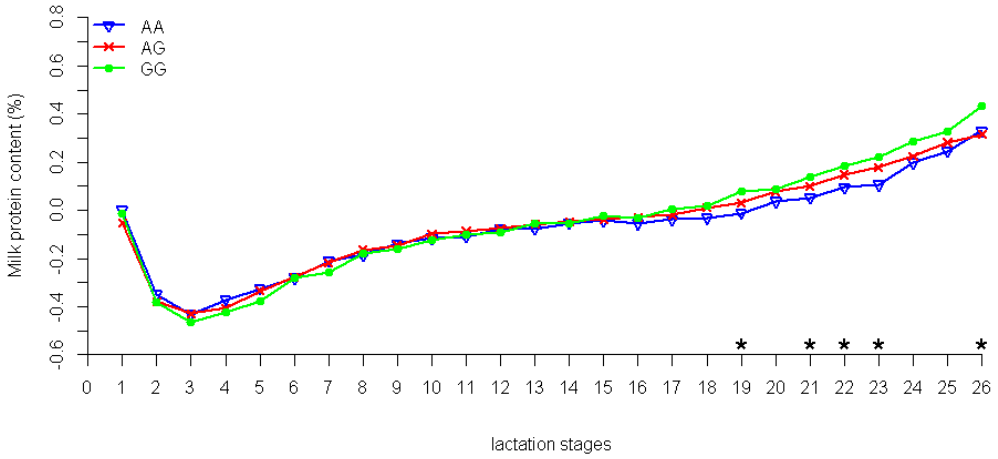


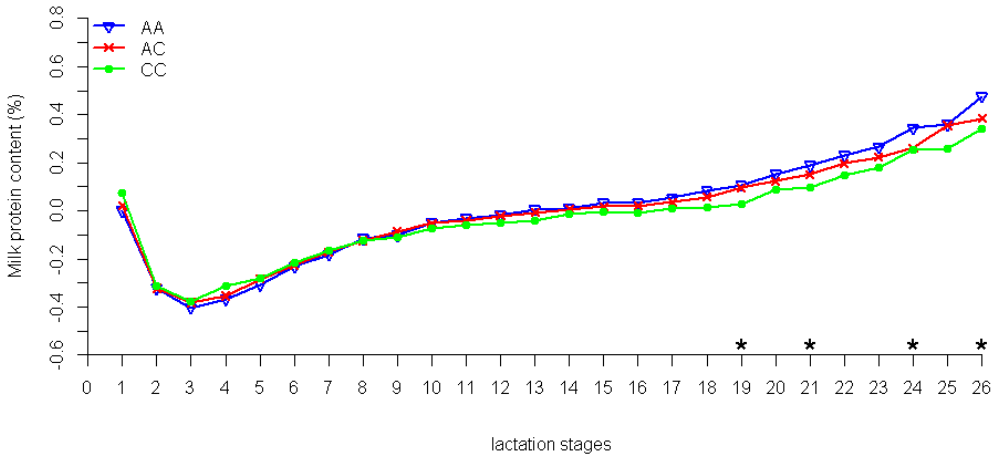
Figure 4. Manhattan plot for SNP by lactation stage interaction on milk protein content. The cut-off value for the y-axis is set at a $-\log_{10}$(P-value) of 30. The horizontal line indicates the genome-wide significance threshold based on permutation ($-\log_{10}$(P-value) = 18.6).

651 A



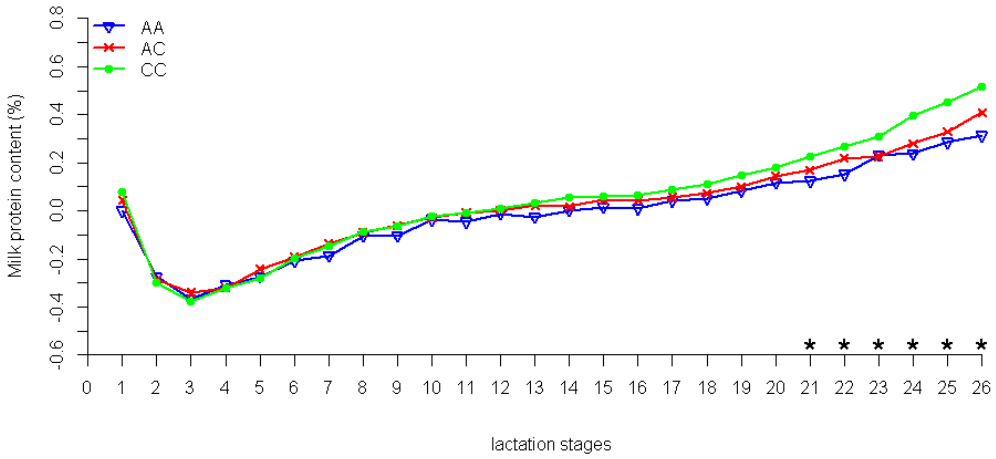
652

653 B



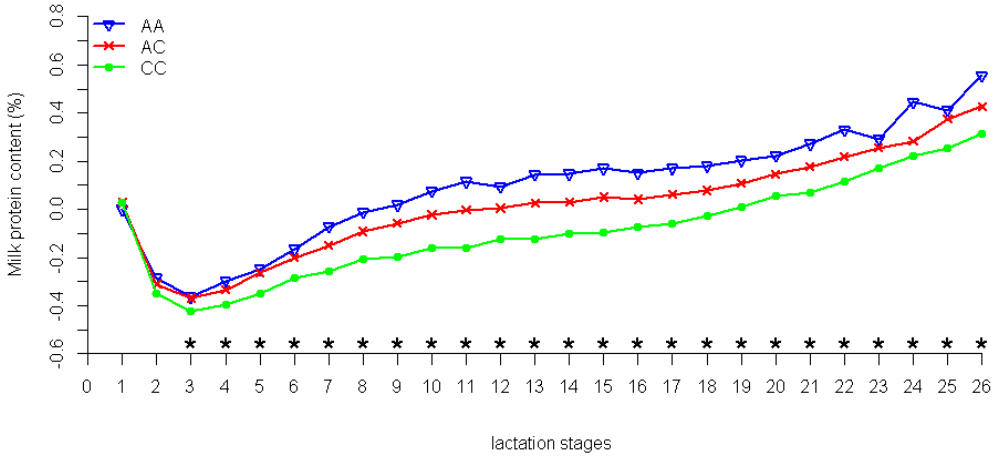
654

655 C



656

657 D



658

659 E

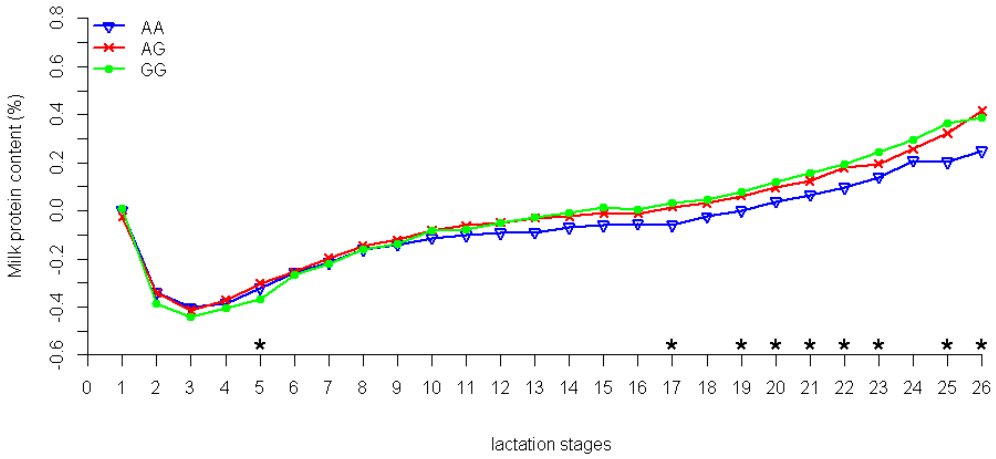


Figure 5. Effects of lead SNP genotypes that show significant SNP by lactation stage interaction during different lactation stages. A: ULGR_rs29011303 on Chromosome 3. B: BTB-02093517 on Chromosome 9. C: ULGR_BTA-68217 on Chromosome 10. D: ULGR_SNP_AJ318490_1b on Chromosome 14. E: ARS-BFGL-NGS-30207 on Chromosome 27. * indicates a significant ($P < 0.001$) difference between any two SNP genotype classes in that specific lactation stage based on a t-test.

667 **Table 1.** The $-\log_{10}$(P-value) of the lead SNP from different Genome-wide association (GWAS) approaches: Separate GWAS for specific lactation

668 stages, GWAS for Wilmink lactation curve parameters, GWAS based on a repeatability model, and GWAS for SNP by lactation stage interaction

SNP name	BTA[1)]	position (bp)[2)]	Lactation stages											Wilmink			repeat[3)]	Interact[4)]
			1-2	3-4	5-6	7-8	9-10	11-12	13-14	15-16	17-18	19-20	21-26	*a*	*b*	*c*		
Significance threshold			5.1	4.6	4.4	4.3	4.1	4.2	4.1	4.0	4.2	4.1	4.6	4.2	5.3	inf	4.0	18.6[5)]
ULGR_AAFC03118332_11420	1	92,196,881	1.5	4.3	3.5	**4.3**	2.2	2.2	1.1	1.3	0.2	1.0	0.2	**4.5**	0.2	1.8	1.7	0.9
ULGR_rs29011303	3	93,216,176	0.5	0.5	0.1	0.4	0.7	0.1	0.0	0.1	0.6	0.0	2.3	2.2	0.4	5.5	0.2	**23.8**
ULGR_AAFC03092560_13525	4	119,336,974	1.9	1.4	3.5	3.1	**4.3**	**4.4**	**4.5**	3.4	2.7	**4.1**	1.9	2.0	0.2	0.4	**4.5**	0.4
ULGR_AAFC03047193_69593	5	123,940,964	0.1	0.5	1.5	1.3	2.2	1.9	**4.7**	3.9	2.7	**4.2**	1.0	1.6	1.2	0.7	2.4	1.6
ARS-BFGL-NGS-27958	6	85,640,056	**7.7**	**15.2**	**14.9**	**8.1**	**11.5**	**10.6**	**7.3**	**8.9**	**11.2**	**7.7**	**5.0**	**10.2**	0.2	0.2	**14.6**	0.3
ARS-BFGL-NGS-103385	7	6,936,993	0.6	0.9	2.1	2.0	3.4	1.7	1.7	2.5	3.6	1.4	2.4	0.3	0.5	1.1	**4.2**	2.6
ARS-BFGL-NGS-23700	8a[6)]	31,495,260	0.3	4.4	1.4	3.4	3.6	2.2	2.6	2.4	2.6	3.5	2.5	2.0	0.6	0.2	**4.0**	6.2
BTB-00348223	8b	54,529,420	0.3	**5.5**	1.2	0.9	1.8	1.0	1.2	0.9	0.4	0.8	0.3	**4.6**	1.3	2.3	1.3	1.0
ULGR_BTA-85063	9a	15,357,200	0.1	1.0	3.1	4.0	3.0	2.6	2.2	1.7	0.5	1.4	1.3	**4.5**	1.8	0.6	2.9	3.1
BTB-02093517	9b	85,934,554	0.3	0.1	0.4	0.0	0.4	0.7	0.5	0.6	1.8	0.7	3.5	1.3	0.7	5.1	0.4	**18.9**
ULGR_BTA-67196	10a	45,610,197	0.5	0.8	1.4	1.4	1.2	2.3	**4.6**	1.3	2.4	2.8	2.0	0.5	0.1	1.2	2.6	4.8
ARS-BFGL-NGS-31031	10b	46,628,033	0.6	0.7	0.6	0.9	1.0	1.0	1.8	2.6	1.2	1.8	**4.6**	1.0	0.1	4.0	2.2	15.5
ULGR_BTA-68217	10b	48,721,829	0.8	0.1	0.6	0.5	0.1	0.2	1.8	0.7	1.2	1.2	2.8	0.8	0.6	4.2	1.4	**27.3**
ULGR_AAFC03042309_74455	10c	51,641,563	0.4	4.5	3.2	2.9	4.0	3.6	**4.4**	3.9	3.6	4.0	2.4	2.4	0.3	0.6	**5.7**	5.9
ARS-BFGL-NGS-74702	11	75,076,326	1.8	2.1	1.3	2.8	3.4	2.4	3.1	1.2	1.2	1.7	2.9	1.9	0.3	0.4	**4.0**	1.4
ULGR_SNP_AJ318490_1b	14a	445,087	0.1	**8.1**	**12.8**	**30.2**	**43.2**	**38.0**	**47.0**	**39.1**	**31.0**	**42.0**	**8.5**	**26.5**	**12.3**	0.5	**33.1**	**74.4**
BTB-00571421	14b	49,132,599	0.2	2.4	3.9	3.5	3.5	2.8	3.9	2.1	2.5	**4.6**	1.5	2.6	0.1	0.2	**4.4**	4.5

ULGR_AAFC03051145_8303	15a	53,245,382	0.6	4.2	**6.1**	**5.0**	**4.2**	2.9	**4.4**	**7.0**	3.3	4.1	1.3	3.4	1.1	0.1	**5.6**	2.4
ULGR_BTA-27068	15b	61,599,974	2.3	2.6	4.2	**5.1**	**4.5**	4.1	3.3	**5.7**	3.8	3.9	1.1	**4.8**	0.9	0.4	**5.3**	0.7
ULGR_BTA-96933	16a	6,593,236	0.4	2.7	1.4	1.6	2.1	2.9	1.3	1.1	3.1	2.4	**7.0**	0.8	1.2	1.2	**4.6**	3.2
ULGR_BTA-121054	16b	29,757,245	0.9	2.5	2.6	2.8	1.4	2.3	0.9	0.5	0.6	0.4	0.3	**4.9**	2.0	3.6	1.5	3.2
ARS-BFGL-NGS-84358	18	39,954,079	0.5	0.0	0.4	1.2	0.7	1.0	1.4	1.2	0.7	1.7	0.6	2.8	**5.3**	1.6	0.5	2.7
ULGR_rs29016098	20	35,900,587	0.7	1.8	**4.9**	**6.1**	**6.3**	**4.9**	**4.3**	2.4	2.2	2.8	2.8	**6.3**	3.2	1.1	**4.9**	2.1
ULGR_BTC-058392	23	51,608,060	0.1	3.5	3.9	3.4	**4.2**	2.7	1.6	2.6	0.4	1.9	0.6	**4.4**	1.7	1.4	2.8	6.3
ULGR_BTA-57368	24	11,476,207	0.3	0.3	0.8	0.6	1.5	2.7	**4.5**	1.8	0.9	2.3	0.8	0.5	0.0	1.0	2.8	1.0
ARS-BFGL-NGS-39823	26	23,530,300	0.0	1.9	**4.8**	3.1	**4.7**	3.1	**4.2**	**4.2**	1.4	2.8	1.9	4.1	1.7	0.2	**4.0**	1.7
ULGR_BTA-40792	26	28,013,558	0.0	1.8	3.2	2.8	3.8	1.5	2.9	2.0	0.2	1.8	0.2	**4.9**	1.4	2.1	1.8	6.2
ARS-BFGL-NGS-30207	27	37,915,598	0.1	0.2	0.8	1.0	0.5	1.2	3.2	1.9	3.0	1.9	2.6	0.3	0.1	2.6	1.6	**28.6**

669 [1)] BTA: Bos taurus autosome.
670 [2)] position of SNP based on Btau 4.0. The sequence of the SNP are in supplementary Table 2.
671 [3)] Repeatability model using all test-day observations.
672 [4)] Repeatability model including a SNP by lactation stage interaction term. Based on a permutation test the 1% genome-wide significance level for
673 the interaction term set at $-\log_{10}$(P-value) = 18.6.
674 [5)] Significance threshold in terms of $-\log_{10}$(P-value). $-\log_{10}$(P-value) in any GWAS approaches were bold if they are greater than corresponding
675 significance threshold.
676 [6)] The different letters for the same chromosome indicate different QTL.